AEC Computing and Applied Mathematics Center

AEC RESEARCH AND DEVELOPMENT REPORT

PHYSICS NYO-7698

THE ITERATIVE SOLUTION OF ELLIPTIC DIFFERENCE EQUATIONS

by

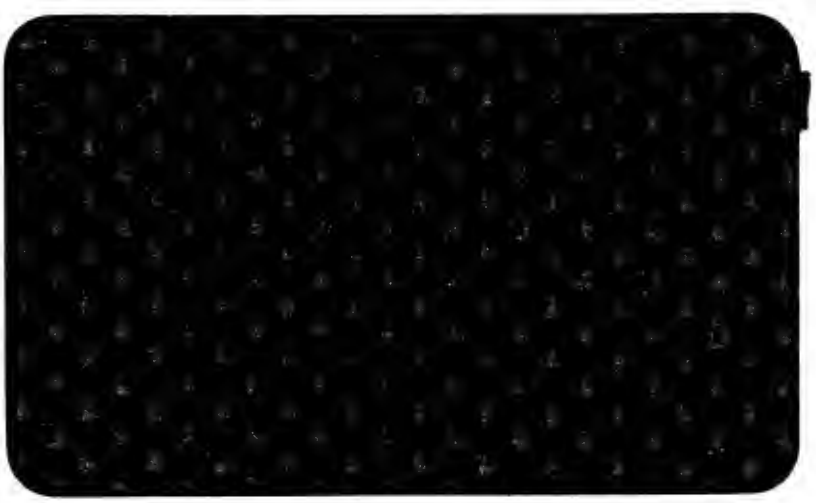
Bernard Friedman

June 1, 1957

Institute of Mathematical Sciences

NEW YORK UNIVERSITY

NEW YORK, NEW YORK



AEC Computing Facility
Institute of Mathematical Sciences
New York University

PHYSICS NYO-7698

THE ITERATIVE SOLUTION OF ELLIPTIC DIFFERENCE EQUATIONS

by

Bernard Friedman

June 1, 1957

CONTRACT NO. AT(30-1)-1480
with the United States Atomic Energy Commission

TABLE OF CONTENTS

ABSTRACT

For a certain class of iteration schemes, it is proved that the method of successive point displacements converges twice as fast as the method of simultaneous point displacements. It is also proved that for such schemes the rate of convergence is unaltered by changes in order in which the iteration is done.

THE ITERATIVE SOLUTION OF ELLIPTIC DIFFERENCE EQUATIONS

I. Introduction

The numerical solution of partial differential equations is usually obtained from the solution of the set of linear equations that are produced by the replacement of the partial derivatives by some finite difference approximations. When the partial differential equation is of the elliptic type, the resulting linear equations are called elliptic difference equations. Since an elliptic equation is associated with a boundary-value problem, the linear equations can not be solved in succession, as is the case with parabolic or hyperbolic partial differential equations, but must all be solved simultaneously; therefore, the problem of solving a set of elliptic difference equations is a special case of the general problem of solving any set of linear equations. However, for elliptic differential equations the coefficient matrix contains many zeros (this will be shown later) and this fact suggests the use of iterative methods for solving the equations.

This note studies some aspects of iterative methods. It is well-known[1] that the rate of convergence of an iterative method depends upon the size of the eigenvalue of maximum absolute value for a matrix which characterizes the iteration method. The use of a simple lemma about determinants (see

Lemma 3) enables us to discuss the eigenvalues for related matrices and thereby to discuss also the rates of convergence of different iteration methods.

The main results of this paper are two. First, a simplified proof of a theorem (see Theorem II) which was first proved in a less general form by Young[2]. This theorem states that in a three-block scheme the method of successive displacements is twice as fast as the method of simultaneous displacements. Second, it is proved (see Theorem III) that the rate of convergence for the method of successive displacements in a three-block scheme is unaffected by the order in which the successive displacements are done. This theorem appears in a slightly different form in an unpublished paper by Heller[3].

II. Fundamentals

Consider a system of p linear equations in p unknowns, $x_1, x_2, \ldots, x_p$. The iterative method for solving these equations is as follows:

Assume, or guess, some values for the unknowns. We call these values the zeroth approximation to the correct values of the unknowns. Using these guessed values, we solve the p equations either individually in some order or by groups to obtain a new set of approximate values for $x_1, \ldots, x_p$. This process by which the new set of approximations was obtained from the original set is repeated* to obtain a second approximation. From the second approximation we obtain a third approximation and so on, hoping that the approximations converge to the true values of the unknowns.

To study the rate of convergence of such an iteration method, we introduce matrices and vectors in a p-dimensional space. Let x be a p-dimensional vector whose components are the correct values of the unknowns. We denote by $x^{(o)}$ the vector whose components are the zeroth approximations, that is, the guessed values, of the unknowns, and by $x^{(n)}$ the vector whose components are the approximations to the unknowns which are obtained after n repetitions of the iteration process. Since the next approximation $x^{(n+1)}$

*It is possible at each step to change the process by which the new approximations are obtained but in this paper we shall not consider this possibility.

is obtained from $x^{(n)}$ by some linear process, it follows that there exists a matrix K and a vector r such that

$$x^{(n+1)} = K x^{(n)} + r \quad . \tag{1}$$

Suppose that for some n we have $x^{(n)} = x$; then if the iteration method is to converge, we must also have $x^{(n+1)} = x$. This implies that

$$x = Kx + r. \tag{2}$$

Consider the error vector $z^{(n)} = x - x^{(n)}$. From (1) and (2) it is clear that

$$z^{(n+1)} = Kz^{(n)} \quad . \tag{3}$$

By a well-known theorem the sequence $z^{(n)}$ will converge to zero, no matter what the zeroth error $z^{(o)}$ is, if and only if the eigenvalues of K are all less than one in absolute value. We define the spectral norm of K as the maximum modulus of the eigenvalues of K. We shall discuss the form of K and its eigenvalues in a few cases.

III. The Method of Simultaneous Point Displacements

The first method we consider is that of simultaneous point displacements. In this method we solve one particular equation for the first unknown x_1 and use the approximate values obtained after n steps for the other unknowns $x_2, \ldots, x_n$ to find the $(n+1)$-th approximation to x_1. To clarify this procedure, let us rearrange the system of linear equations we are solving so that the equation we solve for x_1 is the first equation, for x_2 is the second, and so forth. Suppose the first equation is

$$a_{11}x_1 + a_{12}x_2 + \ldots + a_{1p}x_p = r_1 \quad .$$

We solve for x_1 to get

$$x_1 = -(a_{12}x_2 + \ldots + a_{1p}x_p)/a_{11} + r_1/a_{11} \quad .$$

By substituting the nth approximate values in the right-hand side of this equation we obtain the $(n+1)$-th approximation to x_1.

Similarly, we solve the second equation in the rearranged system of linear equations for x_2 and use the approximate values obtained after n steps for $x_1, x_3, x_4, \ldots, x_n$ to obtain the $(n+1)$-th approximation to x_2. This process is applied to each unknown in turn,

always using the n-th approximate values of all the other unknowns to obtain the (n+1)-th approximation.

Suppose the equations are arranged in the order in which they are solved, that is, arranged so that the k-th equation is solved for x_k. If the linear equations in this arrangement are

$$\sum_{j=1}^{p} a_{kj} x_j = r_k \qquad k = 1, \ldots, p \quad , \tag{4}$$

or, in matrix notation with the obvious meaning for the symbols,

$$Ax = r \quad ,$$

then the (n+1)-th approximation is obtained from the nth by the following formula:

$$x^{(n+1)} = Kx^{(n)} + r' \tag{5}$$

where K is the matrix whose diagonal elements are zero and whose (k,j)-th element is $-a_{kj}/a_{kk}$ and where $r'_k = r_k/a_{kk}$. The convergence of this procedure depends on the eigenvalues of K, that is, on the roots of the equation

$$\det |\lambda I - K| = 0 \quad . \tag{6}$$

If we multiply the k-th row of this determinant by a_{kk} for $k = 1,2, \ldots, p$, the equation will be multiplied by a constant and will therefore have the same roots. However, this multiplication changes the determinant in (6) to the following:

$$(7) \qquad \begin{vmatrix} \lambda a_{11} & a_{12} & \cdots & a_{1p} \\ a_{21} & \lambda a_{22} & \cdots & a_{2p} \\ \cdot & \cdot & \cdots & \cdot \\ a_{p1} & a_{p2} & \cdots & \lambda a_{pp} \end{vmatrix} .$$

We conclude that the zeros of this determinant will determine whether the method of simultaneous point displacements converges or not.

Suppose that instead of solving for the unknowns in the natural order $x_1, x_2, \ldots, x_p$, we decide to solve for them in another order, say in the order $x_{i_1}, x_{i_2}, \ldots, x_{i_p}$, that is, we solve for x_{i_1}, in the first equation, x_{i_2} in the second equation, and so on. Does this change the rate of convergence? In general, it will and the rate of convergence will be determined by the zeros of the following determinant:

$$(8) \qquad \begin{vmatrix} \lambda a_{1i_1} & a_{1i_2} & \cdots & a_{1i_p} \\ a_{2i_1} & \lambda a_{2i_2} & \cdots & a_{2i_p} \\ \cdot & \cdot & \cdots & \cdot \\ a_{pi_1} & a_{pi_2} & \cdots & \lambda a_{pi_p} \end{vmatrix} .$$

Note that (8) is obtained from (7) by changing the order of the columns (without the λ's) from 1, 2, ..., p to $i_1, i_2, \ldots, i_p$ and then inserting λ's along the diagonals.

However, in the case of difference equations arising from differential equations we shall see that it is both natural and logical to solve the i_k equation for x_{i_k}. A system of linear equations will be called distinguished if each equation of the system distinguishes the unknown for which it should be solved. We shall prove the following lemma.

Lemma 1. When the method of simultaneous displacements is applied to a distinguished system of linear equations, the rate of convergence will be the same no matter in what order the equations are solved; it being understood that each equation is solved for the unknown it distinguishes.

Let us so number the equations that the first one distinguishes x_1, the second x_2 and so on. If the equations are solved in the order 1, 2, ..., p, the zeros of the determinant (7) will determine the rate of convergence of the method. Suppose, instead, that the equations are solved in the order $i_1, i_2, \ldots, i_p$. This means rearranging the rows of (7) (without the λ's) so that they are in the order $i_1, i_2, \ldots, i_p$. But, since the i_k-th

equation is to be solved for x_{i_k}, the columns of (7) must also be arranged in the order $i_1, i_2, \ldots, i_p$. The final result is that both the columns and rows of (7) have been subjected to the same permutation $\begin{pmatrix} 1 & 2 \ldots & p \\ i_1 & i_2 \ldots & i_p \end{pmatrix}$.

Under this transformation diagonal elements of (7) will remain on the diagonal. For, by the above transformation the j-th row goes into the j'-th, say, and the j-th column goes into the j'-th; therefore the element a_{jj} goes into the (j',j') place which is also on the diagonal. Since the diagonal elements remain on the diagonal, it is legitimate to consider the above permutation applied to the rows and columns of (7). However a permutation of both the rows and columns of a matrix may be obtained by multiplying the matrix in front by a permutation matrix P and in back by P^{-1}. Such a fore-and-aft multiplication does not change the value of the determinant of the matrix; consequently, the determinant with rows and columns permuted has the same value as the determinant of (7). This proves that the rate of convergence for distinguished equations is not changed by solving the equations in another order.

IV. The Method of Block Displacements

Instead of solving each equation for one unknown, we sometimes prefer to solve a subset of the equations for a certain subset or block of unknowns. This will be called the method of block displacements. If the block consists of just one unknown, we call it a point displacement. Of course, the ideal method would be to solve the complete set of equations for all the unknowns at once. In general, the number of equations is so large that this method is not feasible. However, we may be able to solve for 2, 3, or n unknowns at a time. It is therefore necessary to investigate the rate of convergence of the method of block displacements.

We shall illustrate the discussion of the rate of convergence by considering the case where the block consists of just two unknowns. Going back to the equations (4) we could set up the following iteration scheme:

$$
\begin{aligned}
a_{11}x_1^{(n+1)} + a_{12}x_2^{(n+1)} &= r_1 - \sum_{3}^{p} a_{1j}x_j^{(n)} \\
a_{21}x_1^{(n+1)} + a_{22}x_2^{(n+1)} &= r_2 - \sum_{3}^{p} a_{2j}x_j^{(n)} \\
a_{33}x_3^{(n+1)} + a_{34}x_4^{(n+1)} &= r_3 - a_{31}x_1^{(n)} - a_{32}x_2^{(n)} - \sum_{5}^{p} a_{3j}x_j^{(n)} \\
a_{43}x_4^{(n+1)} + a_{44}x_4^{(n+1)} &= r_4 - a_{41}x_1^{(n)} - a_{42}x_2^{(n)} - \sum_{5}^{p} a_{4j}x_j^{(n)} \\
&\cdots
\end{aligned}
\tag{9}
$$

In this scheme the first two equations would be solved for x_1, x_2, the right-hand side being considered known, the next two for x_3, x_4, and so on.

Introduce the matrices*

$$(10)\qquad N = \begin{pmatrix} a_{11} & a_{12} & 0 & 0 & & \\ a_{21} & a_{22} & 0 & 0 & & 0 \\ 0 & 0 & a_{33} & a_{34} & & \\ 0 & 0 & a_{43} & a_{44} & & \\ & 0 & & & \ddots & \end{pmatrix}, \quad P = \begin{pmatrix} 0 & 0 & a_{13} & a_{14} & \cdots & a_{1p} \\ 0 & 0 & a_{23} & a_{24} & \cdots & a_{2p} \\ a_{31} & a_{32} & 0 & 0 & \cdots & a_{3p} \\ a_{41} & a_{42} & 0 & 0 & \cdots & a_{4p} \\ \cdot & \cdot & \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdots & \cdot \end{pmatrix},$$

then (9) may be written as follows:

$$(11)\qquad Nx^{(n+1)} = r - Px^{(n)} \quad .$$

From this it is easy to see that

$$Nz^{(n+1)} + Pz^{(n)} = 0 \ , \quad \text{or,} \quad z^{(n+1)} = - N^{-1}Pz^{(n)}$$

where $z^{(n)} = x - x^{(n)}$; consequently, the rate of convergence of this method will depend on the eigenvalues of the matrix $-N^{-1}P$, that is, on the zeros of

$$\det |\lambda + N^{-1}P|$$

---

*I use here a suggestion of Dr. H. B. Keller.

or, what is equivalent, the zeros of

$$\det\,|\lambda N + P| = \det N \cdot \det|\lambda + N^{-1}P| \quad .$$

For the case we are considering this determinant has the following form:

$$\det|\lambda N + P| = \begin{vmatrix} \lambda a_{11} & \lambda a_{12} & a_{13} & a_{14} & \cdots & a_{1p} \\ \lambda a_{21} & \lambda a_{22} & a_{23} & a_{24} & \cdots & a_{2p} \\ a_{31} & a_{32} & \lambda a_{33} & \lambda a_{34} & \cdots & a_{3p} \\ a_{41} & a_{42} & \lambda a_{43} & \lambda a_{44} & \cdots & a_{4p} \\ \cdot & \cdot & \cdot & \cdot & \cdots & \cdot \end{vmatrix}$$

In the general case of arbitrary blocks, the rate of convergence will also depend on the zeros of a determinant $|\lambda N + P|$ where N will be the matrix (similar to the left-hand side of (10)) that is inverted to obtain the new approximation, i.e., the $(n+1)$-th, whereas P is the matrix applied to the previous approximation, i.e. the n-th, (similar to the right-hand side of (10)). Note that $N + P$ equals the original matrix of the system (4). From this discussion we may state the following rule:

The rate of convergence of the method of block displacements will depend on the zeros of a certain determinant. This determinant is the same as the original determinant of the system of linear equations except that the coefficients

of each block of unknowns in the equations that are used to determine them are multiplied by λ.

Consider the distinguished set of equations

$$\sum_{j=1}^{p} a_{ij}x_j = r_i \quad , \qquad i = 1, 2, \ldots, p \quad .$$

If $a_{ik} = 0$ for some particular values of i and k, then when the i-th equation is used to find a new approximation to x_i, the value obtained will not depend on the value of x_k. If at least one of the coefficients a_{ik} or a_{ki} is not zero, we shall say the unknowns x_i and x_k are coupled. If both a_{ik} and a_{ki} are zero, we shall say that x_i and x_k are uncoupled. From the previous discussion of block displacements it is easy to show the following:

Lemma 2. If the blocks are composed of uncoupled unknowns, then the method of block displacements is identical with the method of point displacements.

V. Method of Successive Displacements

In the previous methods the components of the (n+1)-th approximation were found by always using the values of the components at the n-th approximation. In the present section we make use of the fact that in practice the components of the (n+1)-th approximation are not all found simultaneously but instead are actually found in succession, one after the other. This suggests that in finding the (n+1)-th approximation to some given component we use the latest available information, that is, the (n+1)-th approximation for any component if it has already been obtained. For example, we may use the following iteration scheme for the equations (4):

$$
\begin{aligned}
a_{11}x_1^{(n+1)} &= r_1 - \sum_{2}^{p} a_{1j}x_j^{(n)} \\
a_{22}x_2^{(n+1)} &= r_2 - a_{21}x_1^{(n+1)} - \sum_{3}^{p} a_{2j}x_j^{(n)} \\
a_{33}x_3^{(n+1)} &= r_3 - a_{31}x_1^{(n+1)} - a_{32}x_2^{(n+1)} - \sum_{4}^{p} a_{3j}x_j^{(n)} \\
&\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\
a_{pp}x_p^{(n+1)} &= r_p - \sum_{1}^{p-1} a_{pj}x_j^{(n+1)} \quad .
\end{aligned}
\tag{12}
$$

This scheme can again be written as

$$Nx^{(n+1)} = r - Px^{(n)}$$

where

$$N = \begin{pmatrix} a_{11} & & & & \\ a_{21} & a_{22} & & & \\ a_{31} & a_{32} & a_{33} & & \\ \cdot & \cdot & \cdot & \cdot & \\ a_{p1} & a_{p2} & a_{p3} & \cdots & a_{pp} \end{pmatrix}$$

and

$$P = \begin{pmatrix} 0 & a_{12} & a_{12} & \cdot & a_{1p} \\ & 0 & a_{23} & \cdot & a_{2p} \\ & & 0 & \cdot & \cdot \\ & & & \cdot & \cdot \\ & & & & 0 \end{pmatrix},$$

the omitted elements being all zero.

We call the iteration method defined by (12) the method of <u>successive point displacements</u>. Just as in the preceding section, the rate of convergence of this method will depend on the zeros of the determinant $|\lambda N + P|$, that is, on the zeros of

$$(13) \qquad \begin{vmatrix} \lambda a_{11} & a_{12} & a_{13} & \cdots & a_{1p} \\ \lambda a_{21} & \lambda a_{22} & a_{23} & \cdots & a_{2p} \\ \lambda a_{31} & \lambda a_{32} & \lambda a_{33} & \cdots & a_{3p} \\ \cdot & \cdot & \cdot & \cdots & \cdot \\ \lambda a_{p1} & \lambda a_{p2} & \lambda a_{p3} & \cdots & \lambda a_{pp} \end{vmatrix}$$

It is also possible to consider a method of <u>successive block displacements</u>. For example, in (9) the iteration scheme would be modified as follows:

$$a_{11}x_1^{(n+1)} + a_{12}x_2^{(n+1)} = r_1 - \sum_3^p a_{1j}x_j^{(n)}$$

$$a_{21}x_1^{(n+1)} + a_{22}x_2^{(n+1)} = r_2 - \sum_3^p a_{2j}x_j^{(n)}$$

$$a_{33}x_3^{(n+1)} + a_{34}x_4^{(n+1)} = r_3 - a_{31}x_1^{(n+1)} - a_{32}x_2^{(n+1)} - \sum_5^p a_{3j}x_j^{(n)}$$

$$a_{43}x_3^{(n+1)} + a_{44}x_4^{(n+1)} = r_4 - a_{41}x_1^{(n+1)} - a_{42}x_2^{(n+1)} - \sum_5^p a_{4j}x_j^{(n)}$$

and so on. The convergence of this scheme would depend on the zeros of the determinant

$$(14) \qquad \begin{vmatrix} \lambda a_{11} & \lambda a_{12} & a_{13} & a_{14} & \cdots & a_{1p} \\ \lambda a_{21} & \lambda a_{23} & a_{23} & a_{24} & \cdots & a_{2p} \\ \lambda a_{31} & \lambda a_{32} & \lambda a_{33} & \lambda a_{34} & \cdots & a_{3p} \\ \lambda a_{41} & \lambda a_{42} & \lambda a_{43} & \lambda a_{44} & \cdots & a_{4p} \\ \cdot & \cdot & \cdot & \cdot & \cdots & \cdot \\ \lambda a_{p1} & \lambda a_{p2} & \lambda a_{p3} & \lambda a_{p4} & \cdots & \lambda a_{pp} \end{vmatrix}$$

A study of λ-determinants such as (13) and (14) show that they have some characteristics in common. The elements of the k-th row of such a determinant are

$\lambda^{\beta_j} a_{kj} (j=1,2,\ldots,p)$ where β_j can be either zero or one. If the value of β_j is zero then $x_j^{(n)}$ should be used in the iteration, whereas if the value of β_j is one then $x_j^{(n+1)}$ should be used. Since the k-th row comes from the k-th equation and since this equation is solved for x_k, we obtain the following iteration scheme:

$$a_{kk} x_k^{(n+1)} = - \sum_{j \neq k} a_{kj} x_j^{(n+\beta_j)} \quad . \tag{15}$$

From this equation we deduce the following characteristics of all λ-determinants:

1) The diagonal elements are always multiplied by λ

2) All those elements whose (n+1)-th approximate values are used in the iteration scheme are multiplied by λ.

VI. Elliptic Difference Equations

When a boundary-value problem for an elliptic partial-differential equation is to be solved numerically, the first step is to introduce a mesh, usually rectangular, into the region on which the boundary values are given. The next step is to replace at each interior mesh point the partial derivatives of the unknown function by some finite difference approximation involving the values of the unknown function at neighboring mesh points. In this way the partial differential equation is replaced by a set of algebraic equations, one equation for each mesh point. These equations are called elliptic difference equations.

For example, consider the problem of solving the potential equation

$$\Delta u = 0$$

in the rectangle $0 \leq x \leq a$, $0 \leq y \leq b$, given the values of u on the boundary of the rectangle. The lines $x = \frac{i}{n+1}a$, $(i=0,1,2,\ldots,n+1)$ and $y = \frac{j}{m+1}b$ $(j=0,1,2,\ldots,m+1)$ will produce a mesh on the rectangle. Put

$$u_{ij} = u\left(\frac{i}{n+1}\, a, \frac{j}{m+1}\, b\right)$$

so that u_{ij} is the value of the unknown function at a mesh point. Using the well-known second difference

approximation to the second derivative, we find that the potential equation is replaced by the following:

$$(16)\qquad \frac{u_{i-1,j}-2u_{i,j}+u_{i+1,j}}{2h_1^2}+\frac{u_{i,j-1}-2u_{i,j}+u_{i,j+1}}{2h_2^2}=0\ ,$$

$$(i=1,2,\ldots n\ ;\quad j=1,2,\ldots m)\ .$$

Here $h_1=\frac{a}{n+1}$ and $h_2=\frac{b}{m+1}$. Note that $u_{0,j}$, $u_{n+1,j}$, $u_{i,0}$ and $u_{i,m+1}$ are known from the given boundary values; consequently, in (16) we have mn equations for the mn unknowns u_{ij} $(i=1,\ldots,n\ ;\ j=1,\ldots,m)$.

Note that equations (16) are distinguished because the equation obtained at any mesh point distinguishes the value of the unknown function at that point. It is theoretically possible to use the equation at one mesh point to solve for the value of the unknown function at another mesh point but in practice this would probably give a method which, if it converges at all, converges very slowly.

Hereafter, we shall consider only elliptic difference equations and therefore only distinguished equations.

VII. A Lemma on Determinants

We shall prove a lemma about determinants that will be used in a later discussion of the question of whether the method of simultaneous displacements or the method of successive displacements converges more rapidly towards the correct solution. Consider the square matrix

$$M = \begin{pmatrix} A_{11} & A_{12} & \cdots & A_{1n} \\ A_{21} & A_{22} & \cdots & A_{2n} \\ \cdot & \cdot & \cdots & \cdot \\ A_{n1} & A_{n2} & \cdots & A_{nn} \end{pmatrix} \tag{17}$$

in which the elements A_{ij} are arbitrary matrices except that the diagonal elements $A_{11}, A_{22}, \ldots, A_{nn}$ are square matrices, not all being necessarily of the same order. Let

$$I_k \ (k=1, 2, \ldots, n)$$

be the identity matrix having the same number of rows and columns as the square matrix A_{kk}. Put

$$Q = \begin{pmatrix} I_1 & & & & \\ & \mu I_2 & & & \\ & & \mu^2 I_3 & & \\ & & & \ddots & \\ & & & & \mu^{n-1} I_n \end{pmatrix}$$

where μ is any non-zero scalar. Note

$$Q^{-1} = \begin{pmatrix} I_1 & & & & \\ & \mu^{-1}I_2 & & & \\ & & \mu^{-2}I_3 & & \\ & & & \ddots & \\ & & & & \mu^{-(n-1)}I_n \end{pmatrix} .$$

We denote the matrix QMQ^{-1} by $\bar{M}$; then it is easy to see that

$$(18)\quad \bar{M} = \begin{pmatrix} A_{11} & \mu^{-1}A_{12} & \mu^{-2}A_{13} & \cdots & \mu^{-(n-1)}A_{1n} \\ \mu A_{21} & A_{22} & \mu^{-1}A_{23} & \cdots & \mu^{-(n-2)}A_{2n} \\ \mu^2 A_{31} & \mu A_{32} & A_{33} & \cdots & \mu^{-(n-3)}A_{3n} \\ \cdot & \cdot & \cdot & \cdots & \cdot \\ \mu^{(n-1)}A_{n-1} & \mu^{n-2}A_{n-2} & \mu^{n-3}A_{n-3} & \cdots & A_{nn} \end{pmatrix}$$

We now state the Lemma which we shall use.

Lemma 3. If M and $\bar{M}$ are the matrices defined in (17) and (18), then $\det M = \det \bar{M}$.

The proof follows from the fact that the determinant of the product of matrices equals the product of the determinants. We have

$$\det \bar{M} = (\det Q)(\det M)(\det Q^{-1})$$

$$= \det M$$

since

$$(\det Q)(\det Q^{-1}) = 1 \quad .$$

VIII. Three Block Schemes and Young's Property A

When the coefficient matrix for a set of elliptic difference equations is considered, it is found that a large number of the elements are zero. The reason for this is that the value of the unknown function at any point depends only upon the values of the function at some set of neighboring points. In other words, the value of the function at any given point is coupled only to the values of the function at some neighboring points. For example in (16) the value of u_{ij} is coupled only to the values of $u_{i-1,j}$, $u_{i+1,j}$, $u_{i,j-1}$, and $u_{i,j+1}$. It is convenient to say in this case that the point (i,j) is coupled to the points $(i-1,j)$, $(i+1,j)$, $(i,j-1)$, and $(i,j+1)$. A convenient way of representing this fact is by the following pattern:

$$\begin{array}{ccc} & x & \\ x & 0 & x \\ & x & \end{array}$$

Instead of considering the coupling between points, we may consider the coupling between blocks of points. For example, in (16) the points on the i-th vertical line are coupled only to the points on the (i-1)-th line and the (i+1)-th line. We shall call this <u>three block coupling</u> since each block (a line) is coupled only to the block preceding it and the block following it. We represent this coupling diagramatically as follows:

x x x x x x x

0 0 0 0 0 0 0

x x x x x x x

An iteration scheme which uses block displacements and in which each block is coupled only to the block before and the block after is called a three-block scheme. For example, in (16) an iteration scheme based on vertical or horizontal line displacements, either simultaneous or successive, will be a three-block scheme. Notice also that the method of simultaneous point displacements in (16) can be considered as a three-block scheme if the lines parallel to a diagonal of the rectangle are taken as blocks. This result follows from these two facts: First, points on the same "diagonal" are uncoupled; consequently, the method of "diagonal" block displacements is the same as the method of point displacements. Second, a diagonal contains all points (i,j) such that $i+j = \text{constant} = k$, say; from (16) it follows that the diagonal defined by the constant k is coupled only to the diagonals defined by $k-1$ and $k+1$. These facts also follow from the geometrical diagram given before. If we combine the pattern for points on the same diagonal, we get:

```
0 x
x 0 x
  x 0 x
    x 0 x
      x 0 x
        x 0
```

This pattern immediately indicates the presence of a three-block scheme.

Three-block schemes are closely related to matrices which possess Young's Property A[2]. A matrix with elements a_{ij} $(i=1,\ldots p;\ j=1,\ldots p)$ is said to have Property A if there exists a decomposition of the integers $1,\ldots,p$ into two non-overlapping non-empty subsets S and T such that if $a_{ij} \neq 0$ then either $i=j$ or one of the integers i and j belongs to S and the other belongs to T. We shall prove the following theorem.

Theorem I. If a matrix M possesses Property A then the method of simultaneous displacements for M is a three-block scheme. Conversely, any matrix for which a three-block scheme exists possesses Property A.

Proof. Assume M possesses property A. We shall say that x_i belongs to S or T if i belongs to S or to T, respectively. If i and i' are non-equal integers which are both in S or in T, then $a_{ii'} = a_{i'i} = 0$ and consequently x_i is uncoupled with $x_{i'}$. If all the unknowns belonging to S are considered as one block and all the unknowns belonging to T are considered as another block, Lemma 2 implies that the method of block displacements for these blocks is the same as the method of point displacements. Since there are only two blocks, it is obviously a special case of a three-block scheme and the first statement of the theorem is proved.

Suppose now that a three-block scheme exists for the matrix M. Let the blocks of unknowns be denoted by A_1, A_2, ..., A_m. We shall define S and T by specifying the unknowns which belong to them. The unknown x_k belongs to S or T accordingly as x_k is in a block having an odd or even subscript. From the three-block property an unknown in a block with an odd (even) subscript is coupled only to the unknowns in a block with an even (odd) subscript. This means that if $a_{ij} \neq 0$, either $i=j$ or x_i and x_j cannot both belong to either S or T. This is Property A; therefore the theorem is proved.

IX. Simultaneous versus Successive Displacements

In the method of simultaneous displacements the (n+1)-th approximation is found by using only the values given previously at the n-th approximation whereas in the method of successive displacements the (n+1)-th approximation is found by using the most recent values, namely, the values at the (n+1)-th approximation if they are available, otherwise, the values at the n-th approximation. It seems plausible to expect the method of successive displacements to converge more rapidly than the method of simultaneous displacements. The following theorem shows that, for three-block schemes, this expectation is justified.

Theorem II. In a three-block scheme the spectral norm for the method of successive displacements is the square of the spectral norm for the method of simultaneous displacements.

This, of course, implies that when the method of simultaneous displacements converges, then the method of succesive displacements converges more rapidly.

The proof follows from Lemma 3 on determinants. In a three-block scheme the determinantal equation for the method of simultaneous displacements will be

$$\det M_a = \begin{vmatrix} \lambda A_1 & B_1 & & & & \\ C_2 & \lambda A_2 & B_2 & & & \\ & C_3 & \lambda A_3 & B_3 & & \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ & & & & C_m & \lambda A_m \end{vmatrix} = f(\lambda) = 0$$

where $A_1, A_2, \ldots, A_m$ are square matrices. For the method of successive displacements the determinantal equation is

$$\det M_i = \begin{vmatrix} \lambda A_1 & B_1 & & & & \\ \lambda C_2 & \lambda A_2 & B_2 & & & \\ & \lambda C_3 & \lambda A_3 & B_3 & & \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ & & & & \lambda C_m & \lambda A_m \end{vmatrix} = 0 \quad .$$

Put $\mu = \lambda^{1/2}$, then by the Lemma on determinants

$$\det M_i = \det \bar{M}_i = \begin{vmatrix} \lambda A_1 & \lambda^{1/2} B_1 & & & \\ \lambda^{1/2} C_2 & \lambda A_2 & \lambda^{1/2} B_2 & & \\ & \lambda^{1/2} C_3 & \lambda A_3 & \lambda^{1/2} B_3 & \\ & \cdot & \cdot & \cdot & \\ & & & \lambda^{1/2} C_m & \lambda A_m \end{vmatrix} \tag{19}$$

$$= \lambda^{p/2} \begin{vmatrix} \lambda^{1/2} A_1 & B_1 & & & \\ C_2 & \lambda^{1/2} A_2 & B_2 & & \\ & C_3 & \lambda^{1/2} A_3 & B_3 & \\ & & \cdot & \cdot & \\ & & & C_m & \lambda^{1/2} A_m \end{vmatrix} = \lambda^{p/2} f(\lambda^{1/2}) = 0 \quad .$$

Suppose the spectral norm for M_a is

$$s = |\lambda_a|$$

then the spectral norm for M_i is

$$t = |\lambda_i|$$

where by (19)

$$\lambda_a = \lambda_i^{1/2} \quad .$$

From this, it follows that $t = s^2$, which was to be proved.

X. Different Orderings

We have shown that for the method of simultaneous displacements the order in which we solved for the unknowns has no effect on the rate of convergence. However, for the method of successive displacements different orderings of the unknowns will give different rates of convergence. In this section we shall consider some cases in which varying the order in which we solve for the unknowns will not affect the rate of convergence.

Let us review the method of successive block displacements where the blocks are labeled 1, 2, ..., n according to the order in which they are solved. (Note that the blocks may reduce to points.) The method is as follows:

An initial guess for the values of all the unknowns in the blocks is made. With the help of this guess, the value of block 1, that is, the values of the unknowns in block 1, is found; the value of block 2 is found by the use of the value of block 1 and the remaining guessed values; the value of block 3 is obtained by using block 2, block 1 and the remaining guessed values; and so on. If G denotes the guessed values, we may represent the order in which successive blocks are found by the following sequence:

(20) G, 1, 2, 3, ..., n, 1, 2, 3, ..., n, 1, 2, 3, ...

This sequence indicates that to find the value of any block

we use the values of all the blocks preceding it and the guessed values of all the blocks which have not been found up to that point. The repetition of the numbers 1, 2, ..., n indicates that after we have found the value of all the blocks from 1 to n, we start over again with block 1 and repeat the process indefinitely.

We prove the following lemma:

Lemma 4. If, in the method of successive block displacements, the method begins with an arbitrary numbered block instead of with block 1 but the succession of blocks is just a cyclical rearrangement of the number 1, 2, ..., n, then the rate of convergence is unaltered.

We present the proof for the case where the blocks are evaluated in the order 2, 3, ..., n, 1 ; the proof for any other cyclical rearrangement would be the same. In this case the method starts with a guess G_1; then the value of block 2 is found by using the guessed values; then the value of block 3 is found by using the value of block 2 and the guess; and so on. The method may be represented by the following sequence:

$$(21) \qquad G_1, 2, 3, \ldots, n, 1, 2, \ldots, n, 1, 2, \ldots$$

Comparing this sequence with (20), we see that they would be exactly the same if G_1 was equal to G in (20) modified by the value found for block 1. Since G and G_1 are

completely arbitrary guesses, it is always possible, given G, to find G_1 and conversely, given G_1, to find G such that the methods represented by (20) and (21) give exactly the same values for the blocks. This proves the lemma.

The question of what effect an arbitrary permutation of the ordering has on the rate of convergence, in general, is still open. However, for a three-block scheme the following theorem provides the answer.

Theorem III*. In a three-block scheme the rate of convergence of the method of successive displacements is the same for all possible orderings of the blocks.

Since all possible permutations of the ordering can be obtained as a product of transpositions, it is sufficient to consider the effect on the rate of convergence of a transposition of two blocks. Suppose, then, that the blocks are solved in the order 1, 3, 2, 4, ..., n. The proof for any other transposition will be similar. The sequence for the "transposed" method is

(22) G_2, 1, 3, 2, 4, ..., n, 1, 3, 2, 4, ..., n, 1, 3, 2, 4, ...

Now, in a three-block scheme any block is coupled only to the block before it, with number one less, and the block after

---

*This result is in an unpublished paper by Dr. J. Heller.

it with number one more. The exceptions to this are blocks 1 and n which are coupled only to blocks 2 and n-1 respectively. In particular, block 3 is coupled only to blocks 2 and 4. This implies that a knowledge of the value of block 1 has no effect on the value of block 3; therefore, we may solve for block 3 before solving for block 1. This means that the method illustrated by (22) is the same as the method illustrated by this sequence:

(23) G_2, 3, 1, 2, 4, ..., n, 3, 1, 2, 4, ..., n, 3, 1, 2, 4, ...

The argument we have used for block 3 may be applied to blocks 4, ..., n. These blocks are not coupled to blocks 1 and 2; consequently, a knowledge of the values of blocks 1 and 2 has no effect on the values of blocks 4, ..., n and conversely; therefore, we may solve for blocks 4, .., n before solving for blocks 1, 2. This means that the method illustrated by (23) is the same as the method illustrated by the following sequence:

(24) G_2, 3, 4, ..., n, 1, 2, 3, 4, ..., n, 1, 2, 3, 4, ...

However, the method illustrated by (24) involves only a cyclical rearrangement of the ordering. By Lemma 4 the rate of convergence for (24) is the same as for (20). Since (22) and (24) are illustrations of essentially the same method, this proves the theorem.

BIBLIOGRAPHY

[1] Frankel, S. P.: Convergence of iterative treatments of partial differential equations, M.T.A.C., vol. 4, pp. 65-75, 1950.

[2] Young, D.: Iterative methods for solving partial difference equations of elliptic type, Trans. A.M.S., vol. 76, pp. 92-111, 1954.

NYU NYO-7698 c.3
Friedman
The iterative solution of
elliptic difference
equations.